To appear in the *Journal of Applied Philosophy* Special Issue on Bias and Context

Last Changes: 11/20/18 DRK
Word Count:, 8716 main body; 12,032 total

Minding the Gap:
Bias, Soft Structures, and the Double Life of Social Norms

By Lacey J. Davidson (davidsl@purdue.edu) and Daniel Kelly (drkelly@purdue.edu, corresponding author)
Purdue University Department of Philosophy[1]

Abstract: We argue that work on norms provides a way to move beyond debates between proponents of individualist and structuralist approaches to bias, oppression, and injustice. We briefly map out the geography of that debate before presenting Charlotte Witt's view, showing how her position, and the normative ascriptivism at its heart, seamlessly connects individuals to the social reality they inhabit. We then describe recent empirical work on the psychology of norms and locate the notions of informal institutions and soft structures with respect to it. Finally, we argue that the empirical resources enrich Witt's ascriptivism, and that the resulting picture shows theorists need not, indeed should not, choose between either the individualist or structuralist camp.

I. Introduction

David Foster Wallace began his now famous Kenyon Commencement Address with the following "didactic little parable-ish" story:

> "There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says 'Morning, boys. How's the water?' And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes 'What the hell is water?'"

Continuing in that same didactic key, the point is that even life's most important features can be so common as to disappear, remain transparent despite their significance, retain their power in part because they are so ubiquitous as to be taken for granted.

The "water" we will try to bring into better view in this paper are the normative contours of the collective social reality that we all move through and the psychological apparatus that allows individuals to navigate it. A primary aim is to describe and demonstrate the relevance of an empirical framework and set of concepts for thinking about **norms**, one that will allow us to better see and understand the ways that norms are both individual and collective, and thus serve to join individuals and communities together. To that end, we will immediately replace the metaphor of water with that of **connective tissue**. Not only is it better suited to our purpose of rendering the typically transparent[2] more visible, but we also welcome the connotations of a fabric that is pliable but tough, whose strength flows from but outstrips that of any of the individual fibers that make it up, and that serves the purpose of binding

different component parts into dynamic wholes. So too, we will argue, with norms, individual psychologies, and the soft but durable social structures that they help constitute.

Our plan of attack is to address a specific debate concerning the best way to approach racism, sexism, and other forms of injustice. The occasion for this debate (or this iteration of it) is the rise in prominence of work implicit bias and the philosophical attention it has received, particularly with respect to morality (see especially Brownstein and Saul 2016). Critics worry that this fixation on implicit cognition is unfortunate, if not outright counterproductive, because it draws attention away from the more significant structural sources of bias and from exploring institution-level mitigation strategies.

We maintain that framing this as a debate, and the strategies as oppositional, is itself counterproductive. Rather, the choice between an individualist approach or a structuralist approach is not a choice anyone needs to make. Those whose primary focus is on institutions and structures will greatly benefit from engaging with the details emerging from the cognitive-scientific work on norms and norm psychology, and those interested in individual hearts and minds will greatly benefit from better appreciating collective level dynamics, especially the emerging and complementary empirical work on norm change, social learning, and cultural evolutionary theory. We briefly map out the geography of the debate in Section II, describing some of the positions taken and the reasoning behind them. In Section III, we use Charlotte Witt's views on gender to help situate our own, showing how her position, and the normative ascriptivism at its heart, provides an initial way to conceptualize how individuals are integrated into the social reality they inhabit. In Section IV we describe recent empirical work suggesting that human minds are equipped with a psychological system dedicated to norm cognition and motivation. Finally, in Section V, we pull the discussions together, laying the ways our position coincides with and departs from Witt's. We flesh out the notions of informal institutions and soft structures and argue that the picture of norm psychology emerging from the human sciences supports a qualified version of Witt's normative ascriptivism. Finally, we revisit the individualist/structuralist debate and point to ways that resources drawn from each approach can be put together into a more integrated, and potentially more fruitful, approach.

II. A Brief Sketch of Conceptual Geography: Individualist and Structuralist Approaches to Bias

Those who wish to understand and change the social world face big questions about how to proceed and how to best direct their efforts. This is exacerbated by the complexity of many social problems. Neither causes nor solutions are straightforward, and it is difficult to know who or what to hold accountable. Simplifying idealizations help, and theorists taking up these challenges can be divided into roughly two camps, individualists and structuralists, according to which initial idealization they favor. Broadly speaking, individualists take primary causes of and solutions to many social injustices to be found at the level of the individual, perhaps within the hearts and minds of those individuals. Structuralists, on the other hand, take the causes of and solutions to social ills like persistent bias, discrimination, and other forms of injustice to lie in the institutions that structure society, i.e. in practices, traditions, social arrangements, laws, and even in the physical distribution of land and goods. While most theorists acknowledge the importance of factors of both types, most often proceed as if one type has primacy over the other.

Over the past several years, philosophical disagreements between individualists and structuralists have re-emerged as research on implicit bias has moved to center stage. Structuralists have worried that concern for implicit bias is far too individual-centric; ridding every person of implicit biases would still fail to e.g. help those living in poverty who lose a job because structural changes to bus routes prevent them from being able to commute. Large scale injustices can occur without any individual intending or personally causing them. Hence, structuralists maintain that being overly focused on individuals can obscure the most significant drivers of social dynamics and so will yield meager resources for affecting deep and abiding social change.[3]

Perhaps the most prominent advocate of **structuralism** within philosophy is Elizabeth Anderson, who argues desegregation is a moral imperative. According to her view, the root and fundamental cause of racial inequities can be linked to *de facto* segregation. Anderson acknowledges other factors, for instance limited public transportation or norms around hiring and job advertisement but sees these causes as structural as well.[4] Similarly, Haslanger argues that the most concerning cases of injustice can be explained solely in terms of structural factors and so without appeal to the mental states of individuals. She imagines three cases focused on, respectively, structures of social life, schemas under which we operate, and policies allocating resources to people in ways that are unequal or inequitable. In her examples, no individual person does anything wrong; rather, the structural factors cause and maintain injustice. Given this premise, it is unsurprising that many structuralists see their approach as superior. They also tacitly assume that if current structures were replaced with more just and unbiased ones, individual changes would result as a 'free' byproduct. Many structuralists nominally acknowledge the necessity of changes to individual hearts and mind, but their theories suggest that, alone, such individual changes will not suffice to bring about robust social change.[5]

Alternatively, many of the most influential proponents of an **individualist approach** are focused on racism and acknowledge that while there are structures that contribute to inequitable outcomes, this is not where the *wrongs* of racism lie. Rather, those wrongs ultimately flow from the attitudes and other mental states of individuals, meaning attention should be paid primarily to those. For example, Garcia holds that racism is affective, in a person's heart; it is a type of moral disregard that individuals have for others *qua* race.[6] Blum offers a variation on this individualist theme, arguing that racism is more cognitive, in a person's head. At bottom it is thinking about and treating other individuals as inferior, or harboring dislike and hostility for others because of their race. Subtle distinctions help support such claims. Blum notes that while there are many manifestations of badness and injustice about which individuals should be concerned, not all should count as racist. Racial *ills* can occur even in the absences of people and beliefs that are genuinely rac*ist*.[7] Both philosophers are often characterized as individualists, because both pursue this general strategy for characterizing racism and embrace its implication that individuals and their mental states ought to be the primary focal point in fighting it.[8]

Since the recent attention given to implicit biases seems to fall squarely in this camp, it is no surprise that structuralist critics have criticized it. They have argued that many types of injustice would still occur even if most people implicit biases were eradicated,[9] and urged

that the (alleged) individualist focus of philosophical discussions of implicit bias draws attention away from the root causes of social injustices.[10]

The prospects for moving beyond these familiar disputes appears promising, though. Madva offers compelling arguments that the push to give priority to structural approaches is, even when coherent, misguided.[11] He points out that since bringing about *structural* change often requires that *individuals* choose to work for it, insisting on a clean separation of the two is misleading. Saul challenges the usefulness of the distinction on similar grounds, supporting her case by pointing out that as an instructor she can make the individual choice to change her grading policy, which would install a structural change that systematically affects all of her students.[12] Another theme running through these discussions is that there are continuous feedback loops of mutual influence between individuals and structures, minds and social worlds.[13]

We see these recent developments as heading in the right direction. So rather than shift away from a focus on empirical psychological research, we will double-down on it. In the next sections, we will explore the crucial role that norms play in connecting individuals and structures, and show how empirical work on norm psychology can deepen the understanding of those connections.

III. Witt On Social Reality

We now turn up the resolution of our discussion on a specific position. The main project of Charlotte Witt's *The Metaphysics of Gender* is to develop an Aristotelian account of gender essentialism[14] – which she calls 'uniessentialism', and which we will unpack below. While pursuing that project, she also occasionally comments on how the conceptual apparatus of her view can help reconceive the aims of feminism, and thus inform efforts to address oppression, bias, and injustice. She sets out her position here in contrast to a familiar type of individual-centric alternative:

> Gender uniessentialism **directs our attention away from individual psychologies**, their conscious and unconscious biases, and "deformed" processes of choice, and toward the social world, its available social roles … I do not mean to dismiss or to criticize important feminist work on deformed preferences or to minimize the role of gender schemas or implicit bias in perpetuating discrimination against women. But **gender uniessentialism points in another direction**, away from a focus on individual psychologies and **toward the social world and its normative structure**, which defines the conditions of agency for women (our bold).[15]

Here we see three of the four main components of our reconstruction of Witt's view. The first is her endorsement of **structuralism** as the best way to understand and address oppression, which we see in her call for more attention to be paid to social structures and her urge that priority be given to changing social roles and the larger structures in which they are found. This goes hand in hand with the second component, which we will call her **relegation of psychology**, seen in her drawing focus away from individuals and their internal psychological processes and choices, and redirecting it rather to the cultural ecology in which individuals find themselves and to the external social structures that shape the

options they can choose between. Witt's emphasis is on feminism and gender, but we believe her strategy here can be effectively generalized to other social groups and injustices.

The third component, her **uniessentialism about gender**, is more complicated. The crucial, and perhaps primary, elements of social reality on Witt's account are social roles and norms. Though she adds nuance to each within her system, the basic ideas are familiar. Social roles are the parts an individual plays in different group contexts, the positions she occupies in various communities and institutional arrangements. Each social role is situated within a network of expectations and guidelines that attach to and shape those parts and positions. To occupy a social role is to be a member of a socially recognized and widely known (within the group) category and thus to be thought of and treated as an instance of that category by members of the community – including, often but not always, in a reflexive way by the occupier of the social role herself. In virtue of this, an individual who occupies a type of social role (e.g. barista, father, director of undergraduate studies) will end up being subject to many of the same sorts of norms and expectations as other individuals who occupy that type of social role.[16]

While we will have more to say about norms below, as a working definition we can construe them as the rules, often unwritten, that organize social life, marking out what behaviors are required, appropriate, permitted, or forbidden for different kinds of people in different circumstances. Witt also provides some artful terminology to draw attention to the kind of influence that norms have on those to whom they apply. Individuals experience the **normative pull** of norms associated with the specific social roles that they occupy, and they experience that pull because they are **responsive to** and **evaluable under** those norms. By "responsive to" Witt means that "the individual's behavior is calibrated in relation to the norm,"[17] and by "evaluable under" she means that "the individual is a candidate for evaluation by others in relation to that norm."[18] Little detail is provided concerning how the calibrating is done or how interpersonal evaluation is converted into behavioral influence, but complaining about this would miss the point. By our lights, the value, especially of the later pair of expressions, is that they explicitly mark that normative influence over an individual's behavior can originate from *within* the individual as well as from *without*, from internal responsiveness and motivation as much as from external social pressure applied in the wake of evaluations made by others in her community.

The pieces are now in place to better grasp Witt's doctrine about the metaphysics of gender, her **uniessentialism**. On this view, gender is a special kind of "mega" social role. Whether one is recognized as a man or woman, for instance, provides an individual with a principle of normative unity, which serves to order and organize all of the other social roles the individual occupies. Gender mega-social roles also thereby prioritize all of the norms an individual is responsive to and evaluable under, helping to determine which will take precedence in scenarios where multiple and conflicting norms might apply.[19]

The need for the principle of unity provided by gender arises, according to Witt, because every individual occupies many social roles of many different types, both at a single time and also over the course of her lifetime. For instance, a single individual might be a daughter, a sister, an ex-ballet dancer, a newly converted Buddhist, a founder of the Younger Womxn's Task Force of Greater Lafayette, a radiologist, and a dues-paying member of the AMA. Some of these social roles require that one choose to enter and identify with the role in order

to be socially recognized as being an instance of the type. Becoming a radiologist requires an explicit and long-term commitment to gaining the relevant license and expertise, and occupation of the position comes with both prestige and rights as well as expectations and obligations, both official and un-. Many other social roles, however, require neither choice nor conscious identification for individuals who occupy them to be subject to their normative pull. For these less voluntary, often more covert social roles, an individual is recognized by others as occupying the role whether or not she realizes it, and whether she would like to be in that role or not. In being so recognized, she is also brought under the pull of the norms associated with the role.[20]

This point leads to the fourth, and for our purposes most interesting, component of Witt's view, her **ascriptivism**. We interpret this as a position about what makes an individual subject to the normative pull of a norm. She throws the idea into relief by contrasting it with voluntarism, the position associated with Kantians like Korsgaard that, roughly, the authority that a norm holds over an individual rests in the individual's voluntary acceptance of the norm and her conscious, deliberate commitment to or endorsement of it.[21] Ascriptivists like Witt, on the other hand, hold that voluntarism cannot account for the full range of norms and social roles. Rather, in many important cases individuals become responsive to and evaluable under norms because those norms are ascribed to them by other members of their community, even if the collective ascription is made without the individuals' knowledge or consent. As Witt points out, individuals do not voluntarily choose what culture they are born into, nor do they choose many of the social roles they come to occupy within it. For those roles, they also do not choose the associated norms whose normative pull they will experience, nor, even, many of the norms they will internalize and apply to themselves. Hence, voluntary acceptance or endorsement need not have much to do with many of the norms individuals will be socially pressured to comply with or feel internally motivated to obey. As she puts it: "The social role is normative for an individual if she occupies a given social position whether or not that individual consciously identifies with or chooses that social position … Rebellion is one way of being responsive to a norm; so is compliance."[22]

One of the many virtues of ascriptivism is that it makes room for the sorts of conflict that can lead to full blown 'rebellion', where an individual is at odds with a norm that others evaluate her by, and chafes under what she experiences as the oppressive social pressure to comply. Indeed, as we will suggest below, ascriptivism, supported with empirical resources, can yield a psychologically plausible story of the internal dimension of such conflicts as well. Ultimately, we are pluralists; voluntarism can capture how some norms shape a person's behavior, but ascriptivism is much better-suited for others. For example, it begins to illuminate how an individual can consciously reject an ascribed norm while nevertheless remaining responsive to it, how she can explicitly and honestly disavow, rebel against, a norm while continuing to feel the force of its normative pull.[23]

IV. Social Norms and the Psychology of Normativity: A View from the Human Behavioral Sciences

Our discussion in this section draws on research from an array of different disciplines and theorists, and so is of necessity compressed.[24] We highlight features of norm psychology that look to be common ground among many of them, but the compiled view is our own, and our presentation of it is tailored to fit the argumentative needs of this paper. Foremost

among those are showing how this view fits with Witt's project, and pointing to ways the resulting picture can broaden our understanding of the psychological microfoundations underlying norms and institutions, the character of the interface between individual minds and soft social structures, and the sources of stability and change in these connective tissues.

The core idea is that individual human minds feature what we will call a **norm system**: a set of fairly functionally integrated psychological mechanisms dedicated to handling information and guiding behavior specifically concerned with norms and the situations they govern. To a first approximation, this system exhibits many of the properties associated with modular,[25] or system 1 cognition.[26] It is a semi-autonomous subsystem of the mind as a whole, whose operation is typically fast, automatic and effortless, and thus often implicit; it can and often does perform its functions outside of awareness and without conscious guidance. From the first-person perspective, the deliverances of this sub-personal system are often intuitive, not seeming to arrive at consciousness as the products of explicit deliberation. Those deliverances can thus be perplexing, experienced as powerfully motivating and perhaps authoritative, yet phenomenologically different from mere urges or personal preferences. They have been described as containing a "puzzling combination of objective and subjective elements."[27]

What does this subsystem of the human mind do, then? One cluster of functions performed by the norm system centers on identifying and internalizing the norms that prevail in the individual's local community and culture. This **acquisition mechanism** draws the individual's attention to salient social interactions happening around her and makes inferences about the rules governing those interactions. This process of social learning is itself largely automatic, as the acquisition mechanism, guided by various constraints and heuristics, allows an individual to intuitively key in on and absorb the normative structure on display in her social environment. The activity of this domain specific learning machinery can be supplemented by verbal instruction provided by elders, teachers, and peers, filling in details about the scope of a rule and the specifics of the behavior it prescribes or proscribes, the parameters of the situations it governs, the types of persons to which it applies, and the form and strength of sanctions appropriate to violations thereof. Norms thus acquired are **internalized** by the individual. After episodes of this form of enculturation, the rule extracted from her social environment is mentally represented in the individual's norm system, along with the relevant information about its parameters.

The other cluster of functions performed by the norm system is aimed at performance and drives the ways an individual acts on those norms she has internalized. This **execution mechanism** is responsible for identifying situations and types of people to which a norm internalized by the individual might apply. It is also responsible for motivating the individual to act in the way specified by the norm, given the particulars of the current situation and the people involved in it. If the rule applies to the individual herself, the execution system will supply motivation for her to obey it, and so behave in a way that conforms to the norm. If the rule is being violated by someone, the execution system will supply motivation for the individual to enforce it, and so behave in ways that sanction the violator and communicate the wrongness of the transgression.

A key claim of this account is that these **normative motivations** produced by the norm system are special: they are intrinsic, non-instrumental, perhaps psychologically primitive.

The injunctions encoded in a norm that has been internalized will influence behavior in ways indicating they are ends in themselves, rather than steps towards something beyond and more primary. People often follow norms not as a means to gain material reward or avoid external punishment, but just because it seems or feels, from a first-person perspective, like the proverbial right thing to do. In these ways, the normative motivations generated by the norm system appear to differ from other motivational states like personal preferences or instrumental desires.[28]

Not only are internalized norms intrinsically motivating, but the motivation is also "two-pronged," both self- and other-directed. On the one hand, it induces the individual to keep her own behavior in conformity with the norm, and, on the other, it induces punitive behaviors (and sometimes generates accompanying reactive attitudes) towards those who fail to do the same. As is the case with other behaviors subserved by system 1 type processes, the norm system can automatically monitor the social environment for signs of what norms apply to the current situation, and, when one is detected, immediately generate the normative motivations that are appropriate to it via a fast, relatively direct link between cue and response. At a collective level, once a rule is represented in the norm systems of sufficiently many members of a community – members who, having internalized the norm, become intrinsically motivated to comply with it themselves and sanction those who don't – the result is a group-level behavioral regularity that is self-stabilizing. The norm is kept in place by each individual members' reliable propensity to comply and punish those who step out of line.

Taking a step back, the basic gist of this way of thinking about individual minds and their component parts will be recognizable to most philosophers, who are by now be familiar with automatic and implicit cognition, as well as the notion of modularity, the general structure of dual process theories, and the controversies surrounding each.[29] We need not take a stand here on how much of the entire mind is modular or how sharp, in general, the divide between system 1 and system 2 processes is. But it will be useful to note a few ways in which the norm system in particular appears to fit and run afoul of the general template. The norm system, and the core features we have drawn attention to, bears many of the marks of being an innate and universal part of human psychological nature.[30] Punishment-stabilized behavioral regularities are ubiquitous in human cultures, where all manner of activities and social arrangements are structured by norms delimiting what is required, appropriate, permitted, and forbidden.[31] Evidence also suggests that behaviors associated with the norm system emerge in development along a fairly regular ontogenetic trajectory, which itself appears consistent across cultures.[32]

Despite the uniform elements in its core structure and operating principles, the norm system is able to support considerable cross-cultural variation in social structures as well. While norms in general are universal, their specific content is not. Human societies exhibit eye-opening diversity in their social arrangements and the particular behavior-guiding rules that govern them. Children tend to acquire those norms prevalent in the culture in which they grow up, and adolescents and adults eventually acquire the specific norms of the roles and positions they come to occupy within their community. These, of course, can vary dramatically from culture to culture, even from community to community and from role to role within a community (as is emphasized by, for instance, theorists of race and gender). These points suggest that as with language acquisition, a person's norm system intuitively

identifies and internalizes those norms it finds in whatever social environment she confronts, picking up on cues from peers, mentors, and others.[33] The resulting picture indicates that however much of the norm system is innately structured, it clearly requires and 'expects' substantial input from experience and social learning to become calibrated enough to successfully perform its functions. Moreover, there is good reason to think this is a domain specific species of social learning whose underlying mechanisms automatically imbue those norms they infer with distinctively two-pronged normative motivations.

Taken together, this picture shows how people become interconnected into cohesive communities, and how adopting the relevant norms allows individuals to get and remain in sync with the groups of which they are members. On the one hand, present in the community that an individual seeks to enter are a number of group level patterns that make up some of the **structures** of that community's social world. Most salient will be broad patterns of punishment-stabilized behavioral regularities. Also salient will be pieces of information that illuminate group boundaries and that indicate how to properly identify and interact with occupants of different kinds of social roles. On the other hand, present in an individual's mind is the universal and relatively automated psychological system that remains alert for the pieces of information that are crucial but distributed across the social environment. This cognitive machinery underpins norm acquisition and performance, and infuses internalized rules with its distinctive kind of two-pronged motivational influence. It thus serves to reliably bring the behavior of the individual into line with whatever norms are acquired, harmonizing their behavior and sensibilities with the social structures that organize the group. The latter (individual level psychological machinery) extracts guidelines via exposure to the former (group level patterns and social structures), allowing the individual to naturally and smoothly enter into the interactions typical of his or her community, and thus contribute to its social dynamics. In this way, social norms form a soft but durable **connective tissue** that binds individuals to groups via cycling loops of mutual influence.

V. Minding the Gap: Soft Structures, Informal Institutions, and Other Dichotomy Busters

Recall the four central components of Witt's view we identified above:

1) **Uniessentialism** about gender
    - Gender provides a social individual with her principle of normative unity
    - This principle takes the form of a mega-social role that orders and organizes all her other social roles and thus all of the norms that apply to her
2) **Structuralism** about understanding and addressing bias, oppression, and injustice
    - More attention needs to be paid to social structures, and more effort directed to changing the normative contours of social reality
    - Social elements like roles, norms, and other structural level entities deserve explanatory and normative priority
3) **Ascriptivism** about normativity
    - Many social roles, and thus norms, are simply ascribed to an individual, making the individual responsive to and evaluable under those norms even in the absence of any decision or consent from the individual him or her self
    - Rejection of pure voluntarist views norms

4) **Relegation** of psychology
    - Too much focus on individuals and the components of individual psychologies is not just misguided but counterproductive
    - Rejection of the idea that social progress can or will be achieved through transformation at the level of individual choice alone

Uniessentialism

We remain neutral on the first component. As mentioned above, uniessentialism about gender is the main thesis of Witt's project, and she marshals these other components in the service of its articulation and defense. Our interests and aims, at least for this paper, lie elsewhere, and so here we take no stand on the uniessentialist thesis, or any other position regarding the metaphysics of gender (though see endnote 23).

Structuralism

Our position regarding the second component is more complicated. On the one hand, we agree with the core idea that more attention to social structures and a richer set of conceptual tools with which to analyze them is required to better understand and fight oppression. On the other, we would reject an extreme or exclusionary version of this structuralist call to arms. We are skeptical of the suggestion that structural accounts can do **all** of the required explanatory work or that they deserve some kind of explanatory **priority**. Rather, we advocate a more ecumenical version, one that holds that whatever else is in the mix, appeal to social structures will be **indispensable** to understanding the sources of injustice and to reducing bias. We further maintain that the empirical framework described in Section IV contains the kinds of conceptual tools that will be indispensable in continuing to bridge the perceived gap between structuralist and individualist approaches.

For instance, it can suggest ways to both broaden and sharpen our understanding of what counts as a structure in the first place. One uncontroversial form of structural change is change in official laws and other explicit rules that make up what we will call formal institutions. Following other theorists in the tradition we are advocating, we understand institutions in general as "the laws, informal rules, and conventions that give durable structure to social interactions in a population".[34] Institutions are structures; they are not ephemeral things, but rather give rise to and support stable collective patterns as opposed to singular or one-off social events. **Formal** institutions are the kinds of political, legal, financial, and other social organizations that are structured by *explicitly* formulated, often *written* down and *publicly* accessible laws, regulations, policies, or decrees. Conditions required for membership in such institutions, and for occupying particular offices and positions within them, are likewise fairly unambiguously stated by the relevant set of governing rules. These also often explicitly specify the scope of the powers and duties attendant to any office, as well as the form and severity of the punishment that will result from the violation of any rules. In formal institutions, these rules are typically not merely shared, but "on the books": gathered together, written down, and codified in constitutions, legal documents, sets of by-laws, and other explicit, official policies.

Informal institutions share some of these characteristics. They too are structures, giving rise to and supporting stable collective patterns rather than unique episodes. Informal institutions also have positions and behavior-guiding rules as core components. However, they differ in important ways as well. Their constituent positions are not offices or jobs to which members are officially appointed, but rather more casual, sometimes covert social roles. As such, the requirements of membership and role occupancy in communities and groups bound together by informal institutions are frequently less explicit or obvious. Moreover, their proprietary rules are not explicit policies and ratified laws, but norms: the generally known but sometimes tacit guidelines, unwritten rules, shared but unofficial standards, implicit arrangements, verbally communicated customs and traditions that organize the social interactions of the community. Penalties for transgressions of these rules are imposed by the community, but they are neither fully standardized nor delivered via any proceduralized apparatus of punishment. Rather enforcement typically takes the form of expressions of disapproval, reputation damaging gossip, ostracism, and other kinds of informal social sanctioning. As Helmke and Levitsky characterize them, informal institutions are "socially shared rules . . . that are created, communicated, and enforced outside of officially sanctioned channels."[35]

Informal institutions are what we think of as **soft structures**. They are obviously not as literally hard as physical or geographical structures. Nor are they as easily visible or particulate as legal structures and formal policies. Still, we maintain that informal institutions, social roles, and norms deserve to be thought of as structure, albeit of a softer and unrulier sort than these contrast classes. As with many features of her physical environment, informal institutions are also actual, sturdy parts of the world external to an individual's skin and skull. They are real and important environmental regularities, patterns of collective behavior that exist (in part) beyond an individual's own head, and to which she needs to be keenly attuned. These structures are also soft, and more fluidly interwoven with the soft selves of the individuals who they help organize and bind together. The norms an individual has internalized are embodied in her own person, existing (in part) as psychological states physically realized in her brain and body, and manifest in her own actions.[36]

Another similarity between soft structures and their harder counterparts is that in many cases neither are chosen by the people who inhabit them. Just as an individual does not decide what geographical location and climate she initially inhabits, or what nation's citizenship she is born into, individuals do not get to hand-pick all of the informal institutions to which they become bound, either. Nevertheless, a person comes to occupy many roles, and so becomes responsive to and evaluable under their attendant norms, whether or not she approves of these structural aspects of her external reality or is happy with her place in them. Informal institutions, social roles, and norms are thus recalcitrant in the way other structures are; most are independent of any one person's preferences and are not immediately responsive to individual decisions and judgments about them. Perhaps most saliently, when she is presented with genuine opportunities to choose, the soft structures that an individual inhabits will affect the range of options and outcomes she has to select from. Thus, they very much shape and constrain her behavior in exactly the ways that advocates of structuralist approaches emphasize.

Of course, soft and hard structures often exist alongside one another, and understanding the complicated relationship between the two is one of the issues we will call more attention

to presently. One final, crucial (dis)analogy is worth noting first, though. A distinctive feature of formal institutions is that, unlike their informal counterparts, they typically include what Hart called **secondary rules**: roughly, rules about how to **change** the rules.[37] Along with first order rules about e.g. what actions are legal or not, or what powers and duties come with a particular office, formal institutions also have on their books official rules that explicitly specify the procedures by which the official rules can be altered. These will say precisely what needs to happen, for instance, for an existing law to be repealed or struck down, for a new by-law to be adopted, for an amendment to be added to a constitution, or for addenda to officially supplement an established policy clarifying how it is to be applied to situations that were previously grey areas. Informal institutions and soft structures conspicuously lack any analogous rules or procedures for changes minor and peripheral or major and globally transformative. Soft structures do, of course, change. But they do not come with an instruction manual that lays out clearly articulated protocols for adding a new norm or removing or altering an extant one. Informal institutions and soft structures do not have any of what we think of a **formal apparatus of change**. This crucial difference raises challenges concerning how to appropriately frame questions about how and why soft structures and informal institutions change and how to most effectively bring about and guide specific, targeted changes. Doing so will most likely require moving beyond simple analogies with formal institutions and a significantly enriched set of conceptual tools.

Because change is still needed. Soft structures, like their harder and more explicit counterparts, can be more or less fair, and norms themselves can be comparatively unjust or oppressive to the individuals who occupy different social roles. As Witt and other structuralists stress, many of the contributors to prejudice lie in features of the world **external** to an oppressed individual's mind. The enemy is certainly **at** the gates, but still largely outside of them. With respect to the individual, these features of informal institutions are supra-personal and relatively rigid. They are not directly under the control or behest of personal level psychological actions like choice, decision, or avowal. But when it comes to unjust and oppressive *norms*, the enemy is very often **inside** the gates as well. People can adopt all kinds of norms, and when they soak up the soft structure of their communities, individuals can **internalize** norms that are unfairly restrictive to occupiers of the very social roles that they occupy. People can thus, in a way we now think of as Foucauldian, come to participate in their own oppression. They can become intrinsically motivated to contribute to systems of social influence that maintain their own subjection and restrict their own options and opportunities, often with the help of comforting stories they strain to tell themselves about why the situation is acceptable.[38] In a dispiritingly real sense, they will thereby become biased against themselves.

Ascriptivism

This brings us to the third component of Witt's view, her ascriptivism, which we endorse.[39] We interpret this position as descriptive, one that primarily concerns the conditions in which an individual comes under the 'normative umbrella' of a norm, and the sources of the normative pull she thereby comes to experience. Supplementing ascriptivism with our account of the norm system provides details about how being subject to normative pull often translates into influence over behavior. As we remarked above, a particularly attractive feature of ascriptivism is that it calls attention to how normative influence over an individual's behavior can come from **within** as well as from **without**. The addition of this

psychological perspective, with its appeal to intrinsic motivation, along with the other automatized properties of the norm system, sheds additional light on how both internal and external sources of influence can be brought to bear in the absence of anyone's conscious choice, and can be effective without anyone's avowal or consent.

Once an individual comes to be recognized as occupying a certain social role, she will have the associated norms ascribed to her, by others and often herself. So, on the assumption that a norm is widely internalized by members of the community, and an individual is widely recognized as being an occupant of the social role that the norm applies to, the individual will be evaluated by reference to the norm and punished by others when she deviates from it. In cases where the norm has also been internalized by that same individual, she will evaluate *herself*, her *own* behavior by reference to it, even if tacitly or subconsciously. The empirical psychological story thus implies that a person's behavior can be **internally responsive to**, and indeed she can even **evaluate herself** according to, norms she has not explicitly accepted, voluntarily committed to, or even consciously acknowledged. Her psychological norm system can generate these effects on its own. Indeed, given the automatic and often implicit nature of the psychological machinery involved, together with the motivational power it imbues to internalized norms, such norms may become part of a person's sense of herself, fueling pre-verbal feelings about who she "really" is, which can be powerful even when somewhat inchoate.[40]

Thus, Witt's ascriptivism, supplemented with the empirical picture, helps make clear how norms live a kind of **double life**. In order to actually act as an effective connective tissue, norms transcend a number of traditional and perhaps intuitive dichotomies. They have both individual and group level properties; they generate both endogenous sub-personal and exogenous supra-personal sources of influence on individual people; they induce behavior both internally, in the form of normative motivations to act, and externally in the form of reliable, socially delivered punishment and reward; they have stabilizing effects on both self and others.

The relegation of psychology

The cost of accepting this expanded and empirically strengthened case for the third component of Witt's view is rejection of her fourth component, the relegation of psychology. This, we maintain, is a price clearly worth paying. Appreciating how individual minds process norms will be just as **indispensable** to understanding the sources of injustice and to reducing bias as appreciating structural factors. Indeed, the two should continue to be integrated. Structuralists' efforts should be united with accounts of normative cognition and motivation, and of the psychological mechanisms that guide how norms are learned and how they spread through populations.

This view of the double life of norms can provide new ways to conceptualize bias and oppressive social structures, pointing to new pathways for more effectively and deliberately changing them. As noted, informal institutions lack a formal apparatus of change, but change nevertheless. The empirical picture suggests that many important factors influencing these decentralized group level dynamics – the cultural evolution of social reality – are properties of individual's norm systems. Luckily there is a burgeoning field that investigates cultural evolution and how it is supported and shaped by different components of human

minds.[41] It suggests that key factors will be **prestige biases**, which nudge cultural learners to adopt norms demonstrated by those held in high regard and seen to have the greatest success and status, and **conformity biases**, which make people more likely to adopt the norm most common among their (actual and aspirational) peers.[42]

Norms and soft structures have tremendous influence on the minds and behaviors of individuals, independently of (and in some ways more directly than) the formal policies or material incentives they live alongside. But norms interact with and influence those formal institutions as well. Another pay off of this work is that it directs us to ask new questions about the way soft structures and formal institutions interact with each other. Many are familiar with examples of failed attempts to force "top down" changes by imposing new formal institutions without considering if and how they will mesh with a community's extant norms and soft structures. For example, at times the US has made forms of aid to areas of Afghanistan conditional on a fixed percentage of leadership roles in newly installed formal institutions (town council-like entities) being filled by women. The condition failed at achieving the intended outcome of initiating local change towards attitudes of gender equity, or pushing entrenched soft structures away from the sexist status quo. Instead, when a community complied with this condition, which was inconsistent with gender norms that were deeply internalized by most Afghans, it appeared to have the effect of rendering the newly formed formal institutions illegitimate in the eyes of much of the population. This interplay seemed to render the appointees less effective, and the imposed formal institutions more likely to be **rejected** than to take root and stabilize on their own.[43]

Another possible outcome is **acquiescence**. In these cases, formal laws are brought to bear on a community or domain of activity that is already organized by a set of norms, and the laws differ from those norms already internalized by the members of the community. Here, when the norms prove too entrenched and the mismatch is judged to be unworkable, the formal institution finally gives up and simply acquiesces, adopting instead the soft structure and enshrining the norms into formal law.[44]

Other attempts are simply **ineffective**. In the case of Batson vs. Kentucky, the US Supreme Court ruled that rejecting potential jurors based only on their race was unconstitutional. There is good reason to believe, however, that the explicit law did nothing to stop the actual practice. It changed neither minds nor norms. Rather, lawyers switched to giving different explicit rationales for rejecting potential jurors, but their decisions remained as racially driven as they were before the ruling. In this case the underlying soft structures and norms concerning race remain **resilient** even in the face of explicit formal legislation designed to uproot them.[45]

These examples help to begin illustrating how this integrationist, norm-based framework can shape future research. It can inform not only top down strategies like those just discussed, but also more bottom up strategies of the sort associated with activists and grassroots movements. For instance, we can distinguish between trying to formulate and promulgate genuinely new norms, on the one hand, and attempts to **normatively reframe** an activity, prying off a norm that was previously governing it, and replacing it with a different, better one. This idea is particularly useful in illustrating how the norms and soft structures of a community influence and interact with formal institutions via the official policy makers who control its formal apparatus of change. In a democracy, norms widely

shared by members of a constituency will influence what policy options are seen as feasible, viable ideas likely to be accepted as laws. Additionally, political actors who occupy formal positions in government institutions will have internalized role-specific norms, some of which will influence how they evaluate policy ideas, which will in turn influence their willingness to act as advocates or opponents for different bills. Climate change activists have recently achieved surprising levels of success using the strategy of normative reframing to advance their agenda.[46]

A more complete account of social change will have to continue wrestling with problems of **scale** as well. Evolutionary theories of culture can provide guidance, as much recent work has explored the different kinds of norms and social dynamics required to support cooperation and collective action in larger societies, rather than just small communities.[47] This work dovetails with an issue that activists often struggle with, that of how to successfully export norms that are prominent and central to their smaller, local subcultures. Such norms are often consciously devised in the face of discussion and refined by rational argumentation by members of those small communities, who are often deeply invested, building their social identity around the issues and values associated with those norms (BLM, women's rights, animal rights, eating habits and food ethics, etc.) These subcultures and smaller communities are of course not completely isolated, but overlap with other subcultures, and are nested within larger groups, including ultimately the overarching society. One possible method for effectively fashioning and promulgating norms is suggested by work on cultural evolution. In brief, competition between these smaller groups could spark micro and mezzo episodes of **cultural group selection** that would generate norms likely to both work and spread, which in this case means being passed on and adopted by more and larger groups.[48]

Or perhaps a different set of tools than those that work to get a norm accepted in a small group will be needed to export it to other groups, let alone to go viral, hit the mainstream, becoming common not just in a particular subculture but in the larger embedding society as well. Activists for many causes seem to intuitively understand the effectiveness of **prestige heuristics** in spreading their norms and values, judging by their efforts to seek out celebrity spokespersons as bearers of their message, especially spokespersons whose proverbial "crossover appeal" allows them to connect with a wider audience. This idea is ripe for development, especially with enriched understanding of norm psychology and how it interacts with extant institutions. The mechanisms and heuristics that guide social learning and norm transmission will be of particular interest, as they may be leveraged by those attempting launch a particular norm from its initial home in a smaller community to the entire society, to the Big Time.

VI. Conclusion

Soft structures are a crucial part of a person's environment. She must locate herself therein and learn how to navigate through and wrestle with and reconcile herself to various facets of those soft structures, many of which are less than fair or ideal. This is no trivial undertaking, and to some extent it is the work of a lifetime. Luckily our minds, or an important part of them, were built to help us do just this.

Witt's position has helped us situate empirical work on norms with respect to recent debates over individualist and structuralist approaches to bias, oppression, and injustice. While we embrace much of it, we reject the relegation of psychology. It can be easy to identify an individualist approach with one that is sensitive to and guided by details of empirical psychological research and a structural approach with one that is indifferent to, if not outright dismissive of, such details. We maintain that this is a mistake. But even on this front we will end with an olive branch. For there is a sense in which we agree with the spirit of Witt's point, that an overly individualist perspective can be distorting. One of many distortions it leads to is a pinching of the imagination, insidiously constricting "psychology" to what can be made sense of through the lens of individual *choice* and personal *decision*. We hold that this **narrowing of psychology** should be rejected as well and hope to have contributed to a corrective trend.

While this paper was developing, we had trouble deciding if we were rejecting the apparent divide between individualism and structuralism as just one more untenable dualism, or trying to transform it from a dichotomy into a trichotomy by offering a third way, or walking between the horns of a dilemma, or doing something else. We have settled on this: individualism and structuralism are useful heuristics, names for very general ways a theorist can first orient herself and form an initial perspective on these kinds of issues. She will choose one depending on her interests, aims, and primary explanatory targets. While obviously helpful, this pair of heuristics also feeds a vision of sharp boundaries, a clean separation between individuals and structures. It also encourages a picture on which there is a gap between concepts and theories that take individuals as their starting point and those that take structures as primary. However, we hold that sustained and serious application of either approach will inevitably (tacitly or otherwise) make assumptions about and help itself to resources typically regarded as belonging to the other.[49] And so what are needed are more and better examples of ways to **mind the gap**: to continue integrating psychological research, conceived of capaciously rather than narrowly, with ideas about structures and institutions, and to continue softening the boundaries between the two. We offer our discussion in this spirit, and hope to have shown that norms, with their double life in the dedicated, sub-personal psychological machinery of individual minds, on the one hand, and in the connective tissue of soft social structures, on the other, serve as an ideal illustrating case.

NOTES

1 Thanks to participants at the Bias in Context conference at the University of Sheffield, the Bias and Meta-philosophy Workshop at KU Leuven, the Seminar on Automaticity and Moral Responsibility at the University of Oslo, the Prejudice: Intersecting Methods and Perspectives Workshop at Washington University, the Panel on Implicit Bias in the Philosophy Classroom arranged by the APA Committee on the Status of Women, and the Symposium on The Logic of Racial Practice: Embodiment, Habitus, and Implicit Bias at the University of Pittsburgh. Special thanks go to Saray Ayala-Lopez, Erin Beeghly, Michael Brownstein, Cameron Evans, Erin Hennes, Jules Holroyd, Alex Madva, Edouard Machery, Uwe Peters, Chris Yeomans, and especially to Charlotte Witt.

2 Between the initial conception and publication of this paper, a fairly dramatic shift has taken place in American (and world) politics, which has been reflected in the ensuing public discourse. One result has been that the importance of norms, and the distinctions we will point to between hard and soft structures, formal and informal institutions, and explicitly articulated policies and laws, on the one hand, and the unwritten rules their operation rests on, on the other, have become much more salient to large portions of even the non-academic world. Our paper aspires to contribute to this trend. For more philosophical

discussion of the recent political scene, see the special issue on Trump and the 2016 Election, *Kennedy Institute of Ethics Journal*, 27, 2.

3 Sally Haslanger, 'Distinguished lecture: Social Structure, narrative and explanation', *Canadian Journal of Philosophy*, 45, 1 (2015): 1-15.

4 Elizabeth Anderson, *The Imperative of Integration.* (Princeton, NJ: Princeton University Press, 2010).

5 See Haslanger op. cit., Sally Haslanger, 'Injustice within systems of coordination and cognition: Comments on Madva for Brains Blog', *The Brains Blog, Symposium on Alex Madva's a Plea for Anti-Anti-Individualism*, retrieved from http://philosophyofbrains.com/2017/03/06/symposium-on-alex-madvas-a-plea-for-anti-anti-individualism.aspx (2017), and Alex Madva, 'A plea for anti-anti-individualism: How oversimple psychology misleads social policy', *Ergo*, 3, 27 (2016): 701-728.

6 Jorge Garcia, 'The heart of racism', *Journal of Social Philosophy*, 27, 1 (1996): 5-45.

7 Lawrence Blum, 'Racism: What it is and what it isn't', *Studies in Philosophy and Education*, 21 (2002): 203-218. Blum's view is in part motivated by the concern that the concept RACIST is being inflated past usefulness, such that it is being applied to anything morally problematic that is connected to race. He argues there needs to be conceptual room for more nuance, other ways of criticizing individuals with respect to their racial attitudes without accusing them of being racist.

8 While there are important differences between theorists within the individualist camp, Garcia appears to occupy the most purely individualist position. He explicitly holds that the structural is reducible to the individual and that changing people's hearts is the key to both eliminating racism and various racial inequities.

9 Saray Ayala and Nadya Vasilyeva, 'Explaining speech injustice: Individualistic vs. structural explanation', in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, eds. R. Dale, C. Jennings, P.P. Maglio, T. Matlock, D.C. Noelle, A. Warlaumount, and J. Yoshimi (2015): 130-135.

10 Ralph Banks and Richard Ford, '(How) Does unconscious bias matter: Law, politics, and racial inequality', *Emory Law Journal*, 58, 5 (2009): 1053-1122, and Ralph Banks and Richard Ford, 'Does unconscious bias matter?', *Poverty & Race*, 20, 5 (2011): 1-2.

11 Madva op. cit.

12 Jennifer Saul, 'Saul Comments on Madva,' retrieved from http://philosophyofbrains.com/2017/03/06/symposium-on-alex-madvas-a-plea-for-anti-anti-individualism.aspx (2017).

13 Saray Ayala, 'Comments on Alex Madva's 'A plea for anti-anti-individualism: How oversimple psychology misleads social policy", *The Brains Blog, Symposium on Alex Madva's a Plea for Anti-Anti-Individualism,* retrieved from http://philosophyofbrains.com/2017/03/06/symposium-on-alex-madvas-a-plea-for-anti-anti-individualism.aspx (2017) and Sally Haslanger, 'Injustice within systems of coordination and cognition: Comments on Madva for Brains Blog', 2017. Members of the two camps occasionally talk past each other as well. Madva helps clarify things with a particularly useful distinction between backward-looking and forward-looking considerations. The backward-looking aspects of a given theory identify the historical causes of bias and injustice, while the forward-looking aspects are concerned with solutions and identifying factors likely to be instrumental in ameliorating bias and redressing injustice. A theorist might hold that backward- and forward-looking aspects are one and the same, but it is perfectly coherent to think they are different. One may hold that the root causes of inequity are unjust social structures, but the way to overcome inequity going forward is by changing individual hearts and minds.

14 Charlotte Witt, *The Metaphysics of Gender* (New York: Oxford University Press, 2011). For more on uniessentialism, see Witt's reply to commentaries by Cudd, Mikkola, and Ásta in Charlotte Witt, '*The Metaphysics of Gender:* Reply to Critics', *Symposia on Gender, Race and Philosophy*, 8, 2, (2012).

15 Witt op. cit. 2011, pages 3-4. Later, she ends her book on a similar note: "To the frequently voiced question, 'Isn't the point of feminism to give women **choices**?' my answer is 'no, not really.' The point of feminism, in my view, is **to retool and reconfigure social structures** so that they do not oppress and exploit women, and the existing networks of social positions and roles (which vary of course from culture to culture) are a prime example of the **social structures that need changing**. The landscape within which women choose and act needs to **change its normative contours**, and then, perhaps, the point of feminism will be adequately captured by the idea of choice" (our bold), Witt op. cit. 2011 p. 132.

16 For an elaboration of this idea in the case of race, see Ron Mallon and Daniel Kelly, 'Making Race Out of Nothing: Psychologically Constrained Social Roles', In *The Oxford Handbook of Philosophy of Social Science*, ed. H. Kincaid (New York: Oxford University Press, 2012): 507 - 529.

17 Witt op. cit. 2011, page 32.

18 Witt op. cit. 2011, page 33.

19 Witt provides an excellent illustration in her discussion of the infamous Philosophy Smoker. The sartorial norms that attach to the generic role *philosopher* are inconsistent with those that attach to *woman,* and balancing them against each other is a delicate and often fraught endeavor: https://aeon.co/essays/would-you-be-the-same-person-if-you-were-a-different-gender

20 For a naturalistic account of social roles and for more on the useful distinction between overt and covert social roles, see Ron Mallon, *The Construction of Human Kinds*, (Oxford: Oxford University Press, 2016).

21 Christine Korsgaard, *Self-constitution: Agency, identity, and integrity*, (Oxford, UK: Oxford University Press 2009).

22 Witt op. cit. 2011, page 43.

23 A host of issues arise concerning this framework and trans and gender queer identities. Perhaps foremost is the challenge that the conjunction of uniessentialism and ascriptivism may entail that if an individual is not socially recognized as an occupant of a particular gender role, and so not ascribed the set of norms attendant to it, then that individual is simply not an instance of that gender. This could have the upshot of disqualifying transgender individuals from being genuine or real members of their gender, or of barring anyone's membership from gender categories that are not widely recognized or ascribed in their community. We would take these to be fatal flaws. Exploring these challenges with the care they deserve is a task that falls beyond the scope of this paper. We have some hopeful thoughts and would especially like to thank Stephanie Kapusta for pushing us to pursue them. But for now we think saying something cursory and so most likely inadequate would be irresponsible, especially given how intricate and fundamental to people's lived identities these matters are.

24 For overviews, see Michelle Gelfand and Joshua Jackson, 'From one mind to many: the emerging science of cultural norms', *Current Opinion in Psychology*, 8 (2016): 175–181 and Maciej Chudek and Joseph Henrich, 'Culture–gene coevolution, norm-psychology and the emergence of human prosociality', *Trends in Cognitive Sciences*, 15, 5 (2011): 218-226. See Daniel Kelly, 'The Psychology of Normative Cognition', *Stanford Encyclopedia of Philosophy* (in preparation) for a more detailed discussion the evidence and outstanding disputes. Finally, see Daniel Kelly and Taylor Davis "Social Norms and Human Normative Psychology," *Social Philosophy and Policy* (In press) for a comparison of the kind of cognitive evolutionary account of social norms we offer in the main text with an alternative self-fulfilling social expectations account offered in Cristina Bicchieri, *The Grammar of Society: the Nature and Dynamics of Social Norms* (New York: Cambridge University Press, 2006) and Cristina Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (New York: Cambridge University Press, 2016).

25 Peter Carruthers, *The Architecture of the Mind*, (New York: Oxford University Press, 2006).

26 Fiery Cushman, Liane Young, and Joshua Greene, 'Multi-system Moral Psychology', in *The Oxford Handbook of Moral Psychology*, ed. J. Doris and The Moral Psychology Research Group (New York: Oxford University Press, 2010), 47-71 and Daniel Kahneman, *Thinking Fast and Slow*, (New York: Farrar, Straus and Giroux, 2011).

27 J. Kyle Stanford, 'The Difference Between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation,' *Behavioral Brain Sciences*, 2018, 1-13.

28 Emotions are likely involved in this story, as suggested by e.g. Paul Rozin, Laura Lowery, Sumio Imada and Jonathan Haidt, 'The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)', *Journal of Personality and Social Psychology*, 76, 4 (1999): 574-586. However, in this paper we remain neutral on the exact nature of the involvement. At this stage, different theorists construe the special character of normative motivations in their own ways, emphasizing different features and describing them in their own terminology. We find the evidence suggesting that internalized norms are in some sense intrinsically motivating to be persuasive, but we also acknowledge that there remains important conceptual work to be done here, clarifying subtly different ways 'intrinsic' might be interpreted, distinguishing other dimensions along which normative motivations might be unique, and separating out and testing more fine-grained hypotheses about their relation to desires and emotion. For some initial work along these lines see Taylor Davis, Erin Hennes, and Leigh Raymond, 'Normative Motivation and Sustainable Behavior: New Insights from an Evolutionary Perspective', *Nature: Sustainability* 1 (2018): 218-224.

29 Jerry Fodor, *The Mind Doesn't Work That Way*, (Cambridge, MA: MIT Press 2001), Kim Sterelny, *Thought in a hostile world*, (New York, Blackwell, 2003), and Celia Heyes, *Cognitive Gadgets: The Cultural Evolution of Thinking*, (Cambridge, MA: Harvard University Press, 2018).

30 Though how much and what parts of the norm system are innate remains to be seen. In Maciek Chudek, Wanying Zhao, and Joseph Henrich, 'Culture-Gene Coevolution, Large-Scale Cooperation, and the Shaping of Human Social Psychology', in *Cooperation and Its Evolution*, eds. Kim Sterelny, Richard Joyce, Brett Calcott,

and Ben Fraser (Cambridge MA: MIT Press, 2013), 425-457., the authors depict the norm system as "a suite of genetically evolved cognitive mechanisms for rapidly perceiving local norms and internalizing them." Our characterization makes no such commitment and is compatible with a range of possibilities. An intriguing one is that the norm system itself may be a **cognitive gadget**, largely culturally inherited rather than innately specified. See Celia Heyes, *Cognitive Gadgets: The Cultural Evolution of Thinking*, (Cambridge, MA: Harvard University Press, 2018) for a general account of relatively automated psychological mechanisms that are themselves socially learned. She calls these cognitive gadgets (for e.g. reading and writing, for playing chess), and contrasts them with the kinds of innate, genetically inherited psychological mechanisms she calls cognitive instincts.

31 See Elinor Ostrom, 'A Behavioral Approach to the Rational Choice Theory of Collective Action', *American Political Science Review,* 92 (1998): 1–22, Elinor Ostrom, 'Collective Action and the Evolution of Social Norms', *Journal of Economic Perspectives*, 14, 3 (2000): 137–158, Peter Richerson and Robert Boyd, *Not By Genes Alone: How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2005) and Joseph Henrich, *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton, NJ: Princeton University Press, 2015).

32 Marco Schmidt, Lucas Butler, Julia Heinz, and Michael Tomasello, 'Young Children See a Single Action and Infer a Social Norm: Promiscuous Normativity in 3-Year-Olds', *Psychological Science*, (2016): 1-11, Larry Nucci, *Education in the moral domain* (Cambridge: Cambridge University Press, 2001), and Michael Tomasello and Amrisha Vaish, 'Origins of Human Cooperation and Morality', *Annual Review of Psychology*, (2013): 231-255.

33 For more discussion of the linguistic analogy that explores commonalities between the psychological mechanisms underlying language, on the one hand, and morality, on the other, see John Mikhail, *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (Cambridge University Press, 2011) and Erica Roedder and Gil Harman, 'Linguistics and Moral Theory', in *The Moral Psychology Handbook*, ed. J. Doris and The Moral Psychology Research Group (New York: Oxford University Press, 2010), 273-296.

34 Robert Boyd and Peter Richerson, 'Gene-Culture Coevolution and the Evolution of Social Institutions', in *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*, eds. C. Engel and W. Singer (Cambridge, MA: The MIT Press, 2008), 306.

35 Gretchen Helmke and Steven Levitsky, *Informal Institutions and Democracy: Lessons from Latin America*, (Baltimore: Johns Hopkins University Press, 2006): 727.

36 See Andy Clark, 'Soft Selves and Ecological Control', in *Distributed Cognition and the Will*, eds. D. Spurrett, D. Ross, H. Kincaid and L. Stephens (Cambridge, MA: The MIT Press, 2007), 110-122 for a discussion of soft selves and the vague and porous boundaries between human minds and their physical and social environments. For more on how soft structures can affect bodies, see Iris Young, *Responsibility for Justice*, (New York: Oxford University Press, 2011). For a clear analysis of the different views that get discussed under the name of embodied cognition in philosophy of mind and the cognitive sciences, see Lawrence Shapiro, *Embodied Cognition*, (New York: Routledge Press, 2010).

37 H. L. A. Hart and Leslie Green, *The Concept of Law, 3rd Edition, Clarendon Law Series*, (Oxford: Oxford University Press, 2012).

38 See John Doris *Talking to Ourselves: Reflection, Ignorance, and Agency*, (Oxford: Oxford University Press, 2015) for discussion that highlights the social functions of confabulation and post-hoc rationalization with a polemic thrust concerning theories of agency, John Jost, Aaron Kay, and Hulda Thorisdottir, *Social and psychological bases of ideology and system justification*, (New York, NY: Oxford University Press, 2009) for work on the psychology of ideology and system justification, and Serene Khader, *Adaptive Preferences and Women's Empowerment*, (New York, NY: Oxford University Press, 2011) on adaptive preferences.

39 As noted above, we are pluralists with respect the psychology of norms, so we may disagree with Witt about the proper scope of the doctrine. We expressed skepticism in Section III that voluntarism *alone* could account for the full range of norms and social roles, and we are also skeptical that ascriptivism *alone* can account for the full range of norms and social roles. A full theory of human norm psychology will need to be at least two-tiered, able to accommodate both automatically internalized norms and consciously avowed norms, and explain the similarities and differences in how each kind is acquired and cognized and perhaps most importantly, how each motivates behavior. We are highlighting ascribed social roles and internalized norms here, but see Daniel Kelly 'How to Adopt a Norm,' in preparation for initial discussion of a more pluralist approach.

40 See Daniel Kelly and Nicolae Morar 'I Eat, Therefore I Am: Disgust and the Intersection of Food and Identity', to appear in *The Oxford Handbook of Food Ethics*, eds. A. Barnhill, M. Budolfson and T. Doggett

(Oxford University Press, 2018), 637-657 for some discussion of norms, identities, and food that proceeds along these lines, and some speculation about its potential ethical significance.

41 Joseph Henrich, *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*, (Princeton, NJ: Princeton University Press, 2015), Oliver Morin, *How traditions live and die*, (Oxford: Oxford University Press, 2016), Daniel Kelly and Patrick Hoburg, 'A Tale of Two Processes: On Joseph Henrich's *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*,' *Philosophical Psychology*, 30, 6 (2017): 832-848, and Kim Sterelny, 'Cultural evolution in California and Paris', *Studies in History and Philosophy of Biological and Biomedical Sciences*, 62 (2017): 42-50.

42 See Daniel Kelly and Taylor Davis, 'Social Norms and Human Normative Psychology', *Social Philosophy and Policy* (in press) for citations and more detailed discussion.

43 For some discussion, see Aarya Nijat and Jennifer Murtazashvili, 'Women's Leadership Roles in Afghanistan', *United States Institute of Peace: Special Report 380*, September 2015, retrieved from https://www.usip.org/sites/default/files/SR380-Women-s-Leadership-Roles-in-Afghanistan.pdf. For a discussion of these kinds of challenges in slightly different terms, see Ryan Muldoon, 'Perspectives, norms, and agency', *Social Philosophy and Policy*, 34, 2 (2017): 260-276. See pages 328-331 of Henrich op. cit. 2015 for a tongue-only-slightly-in-cheek description of some of the failures of Operation Iraqi Freedom viewed through the lens of interactions between cultural norms and formal institutions.

44 The 1872 Mining Act and "law of the camp" provide an example. For discussion see William Colby, 'Mining Law in Recent Years', *California Law Review*, 33, 3 (1945): 368-387.

45 See the "Object Away" episode of the *More Perfect* podcast for more details: http://www.wnyc.org/story/object-anyway/

46 See Leigh Raymond, S. Laurel Weldon, Daniel Kelly, Ximena Arriaga, and Ann Marie Clark, 'Making Change: Norms and Informal Institutions as Solutions to "Intractable" Global Problems', *Political Research Quarterly*, 67, 1 (2013): 197-211, and Leigh Raymond *Reclaiming the Atmospheric Commons*, (Cambridge, MA: The MIT Press, 2016) for a full treatment of a case of normative reframing that was, unexpectedly, largely successful. Many working on climate change and other environmental issues are actively exploring norm-based strategies as well, suggesting for instance that we abandon the frames that current dominating discussion, e.g. harm, fairness, welfare, and self-interest, and normatively reframe them in terms of purity and sanctity, e.g. Joshua Rottman, Deborah Keleman, and Liane Young, 'Hindering Harm and Preserving Purity: How Can Moral Psychology Save the Planet?', *Philosophy Compass*, (2014): 1-11.

47 Peter Richerson and Joseph Henrich, 'Tribal Social Instincts and the Cultural Evolution of Institutions to Solve Collective Action Problems', *Cliodynamics*: *The Journal of Theoretical and Mathematical History*, 3, 1 (2012): 38-80, Kim Sterelny, 'Cooperation, Culture, and Conflict', *The British Journal for the Philosophy of Science*, 67, 1 (2014): 1-31, and Robert Boyd, *A Different Kind of Animal: How Culture Made Humans Exceptionally Adaptable and Cooperative*, (Princeton NJ, Princeton University Press, 2017).

48 For discussion of actual examples of this process operating on sustainability norms, see Tim Waring, Sandra Goff, Paul Smaldino, 'The coevolution of economic institutions and sustainable consumption via cultural group selection', *Ecological Economics* 131, (2017): 524–532, and Michelle Kline, Tim Waring, and Jonathan Salerno, 'Designing Cultural Multilevel Selection Research for Sustainability Science', *Sustainability Science* 13, 1 (2018): 9-19.

49 As noted in Section II, we are not the first to make this kind of point, and take ourselves to be fellow travelers with Alex Madva, Jennifer Saul, and others.